

MULTIPLE CRITERIA DECISION ANALYSIS FOR HEALTH TECHNOLOGY ASSESSMENT

REPORT BY THE DECISION SUPPORT UNIT

February 2011

Praveen Thokala

School of Health and Related Research, University of Sheffield, UK

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734

E-mail dsuadmin@sheffield.ac.uk

Website www.nicedsu.org.uk

Acknowledgements

The author is grateful for financial support from the NICE Decision Support Unit (DSU). NICE had no responsibility or control over the content of this paper.

EXECUTIVE SUMMARY

This paper aims to look at the applicability of multi criteria decision analysis (MCDA) for health technology assessment.

MCDA is aimed at supporting decision makers faced with evaluating alternatives, taking into account multiple, and often conflictive, criteria. This manuscript begins with a critical review of state-of-the-art methods for incorporating multiple criteria in health technology assessment (HTA). An overview of MCDA is provided and is compared against the current NICE (National Institute for Health and Clinical Excellence) health technology appraisal process. A generic MCDA modelling approach is described and the most common types of MCDA models are detailed. The different MCDA modelling approaches are applied to a hypothetical case study. Finally, the issues that need to be considered for the application of MCDA in HTA are examined along with recommendations for future research.

Most of the proposed MCDA approaches in literature use the same technique (weighted sum approach) which may lead to the researchers/health professionals assuming that it is the only relevant MCDA method. MCDA does not just stop at simple weighting and scoring; more flexible approaches are available that appear to be more relevant to the NICE appraisal process and value based pricing (VBP).

There is a semblance between main MCDA modelling approaches and other techniques (such as programme budgeting and marginal analysis [PBMA], VBP and NICE recommended table of the summary characteristics). However, there are general practical issues that might arise from using an MCDA approach in the HTA process and it is suggested that appropriate care needs to be taken to address the issues identified in order to ensure the success of MCDA techniques in the appraisal process.

CONTENTS

1. INTRODUCTION..... 5

2. INCORPORATING MULTIPLE CRITERIA IN HEALTH TECHNOLOGY APPRAISALS..... 5

3. MCDA vs. NICE APPRAISAL PROCESS..... 7

4. MCDA MODELLING..... 9

5. (A HYPOTHETICAL) CASE STUDY..... 10

 5.1 VALUE MEASUREMENT MODELS..... 11

 5.2 OUTRANKING APPROACH..... 13

 5.3 GOAL PROGRAMMING..... 16

6. GENERIC ISSUES WITH USING MCDA..... 19

7. CONCLUSIONS..... 22

8. REFERENCES..... 23

TABLES AND FIGURES

Table 1 Characteristics of the drugs in the appraisal process..... 10

Table 2 Performance levels of drugs..... 12

Table 3a Performance scores of drugs..... 14

Table 3b Outranking relations and weights..... 14

Table 4 Attributes of drugs, the goals and weights against different criteria..... 17

Figure 1 MCDA and NICE technology appraisal process..... 8

1. INTRODUCTION

In the UK, the National Institute for Health and Clinical Excellence (NICE) makes recommendations to NHS after assessing new and existing medical technologies. The current practice of NICE health technology appraisals is based on the incremental cost-effectiveness ratio (ICER) i.e. the incremental cost per quality adjusted life year (QALY) gained by recipients of treatment. Even though NICE considers other things (e.g. severity, life saving, etc) along with ICERs, there is concern that this approach may fail to capture other important sources of value.¹⁻³ In recognition of this issue, NICE commissioned Professor Sir Ian Kennedy to carry out a study on the relationship between innovation and the value of the technologies.⁴ Also, recent developments such as the patient protection and affordable care act (ACA) in America⁵ and the Department of Health's decision to use value based pricing⁶ indicate a paradigm shift towards using other criteria along with the traditional cost-effectiveness analysis. Multi-criteria decision analysis can support decision makers faced with evaluating alternatives taking into multiple, and often conflictive, criteria in an explicit manner⁷. In fact, a number of manufacturers recommended the use of MCDA (in their submissions to Professor Sir Ian Kennedy) but recognised that further research is needed before their implementation in the health technology appraisal process. This paper looks at the applicability of using MCDA techniques for evaluating health technologies.

2. INCORPORATING MULTIPLE CRITERIA IN HEALTH TECHNOLOGY APPRAISALS

Decision makers so far have been considering cost-per-QALY ratios alongside other criteria, such as equity and fairness, and prioritisation of interventions for vulnerable populations, in a deliberative manner. An integrated ICER which includes other sources of value has been proposed to allow explicit incorporation of other criteria, such as societal preferences, disease severity, equity and benefits to caregivers, in the existing ICER framework. Societal preferences relating to distributional justice have been captured from surveys and included in the ICER calculations.^{8,9} Explicit incorporation of equity in calculating ICERs for health technology assessments has also been considered,^{2,10,11} but a need for further research has been identified.¹²

A hybrid method which supplements the current ICER evaluation for NICE with a comprehensive benefits and value (CBV) review has also been proposed.^{1,13} This approach attempts to capture the sources of value not systematically considered at the present (such as innovation, societal benefit, disease severity, unmet need, patient compliance and related benefits) by using different ICER thresholds for different CBV scores.¹³

Multi-criteria decision analysis has been extensively used to inform healthcare decisions,¹⁴⁻¹⁶ setting priorities for HTAs¹⁷ and other governmental issues.^{18,19} The benefit-risk assessment of medicines, based on multiple benefit and risk criteria including the trade-offs between the benefits and the risks, was performed using MCDA.^{20,21} MCDA techniques have also been used for shared decision making between patients and doctors in the evaluation and selection of therapies, treatments, and health care technologies.^{22,23} These MCDA techniques were said to identify and include the personal preferences of the patient but the complexity of the MCDA models and the time taken to complete the model were mentioned as disadvantages.^{24,25} Program budgeting and marginal analysis (PBMA),²⁶⁻²⁸ used for reallocation of scarce healthcare resources, is also based on MCDA methodologies. This method has received some attention in health sector,^{29,30} but its success has been limited due to the complexity of the approach, large data requirements and organisational barriers.^{31,32}

Despite the use of MCDA in other health streams, it is only recently that there have been studies that advocate the use of MCDA for HTA. A framework utilising a value matrix was developed to include quantifiable components that are currently considered in health decision-making to promote transparent and efficient healthcare decision-making.³³ This framework was also linked to a qualitative assessment including six ethical and health system-related components of decision to provide a tool for combining HTA, MCDA, values and ethics³⁴. A Health England Leading Prioritisation (H.E.L.P) study also used MCDA to prioritise investment in preventative health interventions.³⁵ There have also been online publications (knols) that state that the future of HTA is MCDA, by citing a simple case study.³⁶

However, most of the proposed MCDA approaches use the same technique (weighted sum approach described in section 5.1) which may lead to the researchers/health professionals assuming that it is the only relevant MCDA method. This paper attempts to provide an overview

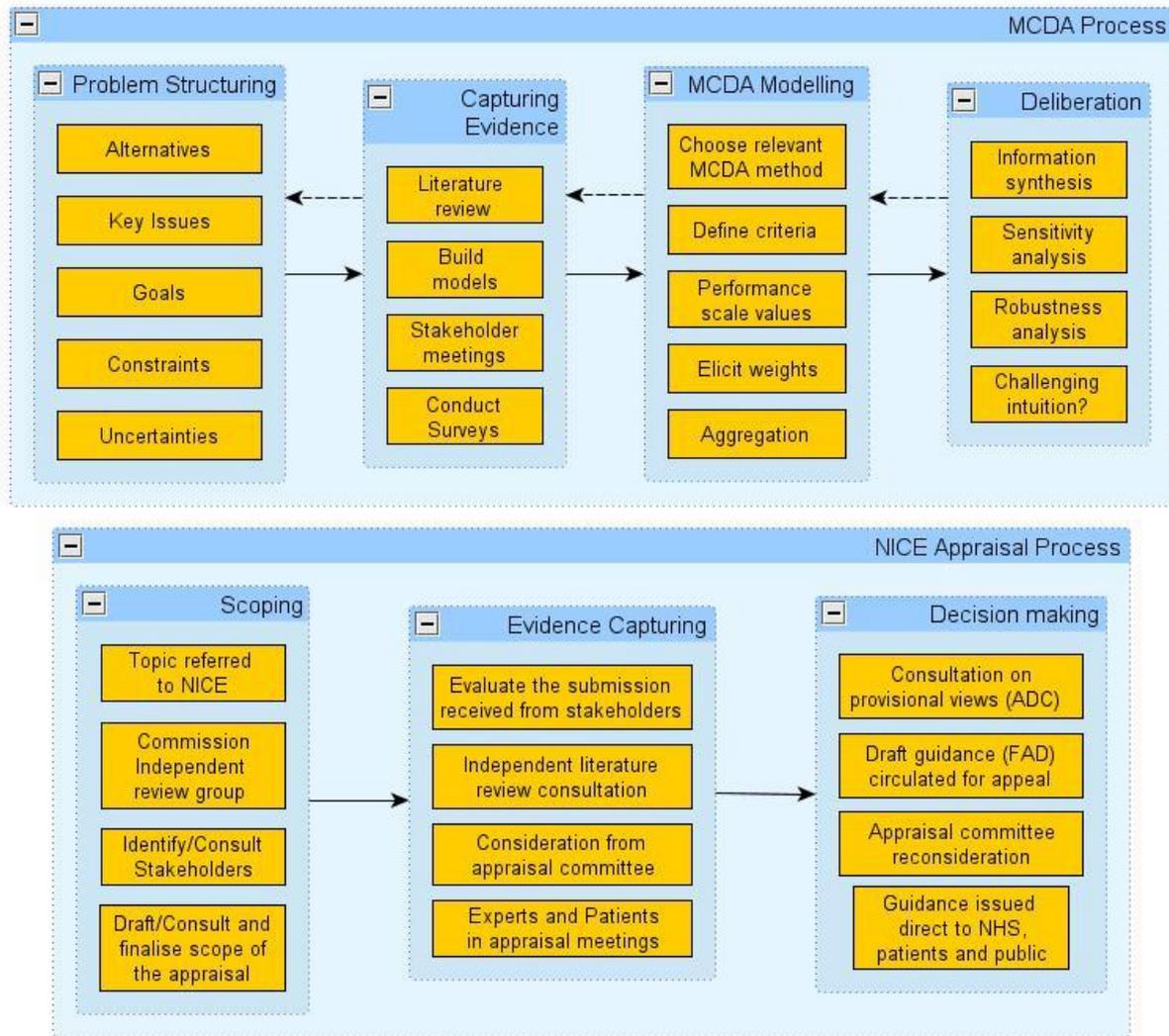
of all the main MCDA methods available and the issues with their implementation in a technology appraisal process.

3. MCDA vs. NICE APPRAISAL PROCESS

MCDA is aimed at supporting decision makers faced with evaluating alternatives taking into account multiple, and often conflictive, criteria. The MCDA process consists of the following phases: problem identification and structuring, model building and use; and the development of action plans. MCDA modelling consists of four key elements: (1) the alternatives to be appraised, (2) the criteria (or attributes) against which the alternatives are appraised, (3) scores that reflect the value of an alternative's expected performance on the criteria, and (4) criteria weights that measure the relative values of each criterion as compared to others.

The MCDA process is compared with the current NICE technology appraisal process as shown in Figure 1. The first two steps of the MCDA process i.e. identifying alternatives and criteria, is known as problem structuring; this is usually achieved by decision conferencing,^{37,38} which involves the meeting of all the relevant stakeholders. The current NICE approach includes this problem-structuring process during the "scoping" stage to set the pre-defined options (treatments, drugs, etc) and the key outcomes relevant for the appraisal process. The criteria for NICE appraisals are defined in the methods guide, not separately for each appraisal, but the scoping process allows identification of other key issues (such as disease specific outcomes, etc). Thus, there is not much difference between the current NICE scoping approach and the first two steps of the MCDA process.

Figure 1 MCDA and NICE technology appraisal process



It is in the decision making stage that the MCDA and the current NICE appraisal processes differ. In the NICE approach, the evidence regarding the alternatives is captured and evaluated in a deliberative manner using ICER and other criteria. In the MCDA approach, this evidence needs to be quantified and input into mathematical models to identify the best alternative(s). The manner in which these models are built separates the different MCDA techniques. Most of the literature on MCDA focuses on evaluation of a “given problem with a well defined set of alternatives and criteria” i.e. building the models, and this will also be the focus of this paper. This paper does not argue for or against the need for a formal mathematical approach in the

NICE technology appraisal process, it just provides an overview of the different mathematical approaches of MCDA.

4. MCDA MODELLING

MCDA modelling comprises of using formal, transparent, mathematical approaches to measure the overall performance of the alternatives on multiple criteria. This is achieved by a) measuring the desirability of achieving different levels of performance in each criteria and b) combining these preferences across individual criteria allowing for inter-criteria comparisons. The manner in which this is modelled separates the different MCDA techniques.

The different MCDA approaches can be broadly classified into three categories⁷:

Value measurement models: The degree to which one decision option is preferred to another is represented by constructing and comparing numerical scores (overall value). The scores are developed for each individual criterion initially and aggregated into higher level value models. Almost everyone who has suggested using MCDA methodology for health technology assessment suggested this approach,^{13,33,34,36} however this approach is not without its constraints as explained in section 5.1. Program budgeting and marginal analysis (PBMA)^{26,28,29} and analytic hierarchy process (AHP),^{39,40} another widely used MCDA technique, are also based on this value measurement modelling approach.

Outranking models: The alternatives are compared pair wise, initially in terms of each criterion, in order to assert the extent of preference for one over the other for that criterion. The preference information across all criteria is aggregated to establish the strength of evidence favouring selection of one alternative over other. This approach is not widely used but could also be an appropriate alternative for MCDA in HTA as it is based on direct comparison of the key characteristics of the drugs/treatments; this is in line with the new NICE recommended table of the summary characteristics in the ERG report.⁴¹

Goal, aspiration or reference level models: This approach involves derivation of the alternative(s) which are closest to achieving the pre-defined desirable (or satisfactory) levels of achievement for each criterion.⁴² Value based pricing,^{43,44} used to set the prices of

drugs/treatments such that the ICER is under the relevant cost-effectiveness threshold, could be implemented using this MCDA approach, provided the definition of “value” is clearly identified by the health organisations such as NICE or NHS.

5. (A HYPOTHETICAL) CASE STUDY

A hypothetical NICE technology appraisal process is considered with a recommendation needed to be made between two new drugs A and B. The characteristics of each of the drugs when compared against the best standard care are shown in Table 1. The cost effectiveness (C/E) of the drugs is measured using net benefit (NB) calculated assuming a willingness to pay (WTP) of £20,000/QALY; if C/E is used to make the decision, drug B would be recommended over drug A. However, MCDA could also be used to compare the drugs; the three different MCDA approaches mentioned in section 4 have been applied to this case study and are described below. It should be noted that the criteria specified here are for demonstration purposes only. The aim of this study is not to enter into a debate on what criteria should be used and their definitions but to compare different MCDA techniques for incorporating multiple criteria into the decision process, once the relevant criteria are identified.

Table 1 Characteristics of the drugs in the appraisal process

	Drug A $z_i(a)$	Drug B $z_i(b)$
C/E (in terms of NB)	£15,850	£25,600
Equity (%)	0.14	0.08
Innovation	Innovative	Less Innovative
Patient compliance (%)	0.93	0.85
Quality of evidence	Good	Good

Before the MCDA models can be developed, the performance of the alternatives (drugs A and B) against the specified criteria needs to be measured in an objective manner. In order to achieve this, measures (or scales) which can describe the desirability of achieving different levels of

performance for each criterion need to be identified. For some criteria, such as patient compliance, where preferences are linearly related to the attribute's value, the attribute value z_i can substitute for the performance on the criterion. However, in most cases, this needs to be modelled as there is rarely such a simple linear relationship between attribute values (z_i) and preferences. In such cases, a scale needs to be constructed to represent the performance of alternatives; it should be noted that choosing the scales (usually ordinal or ordered-categorical scales) to model these performance measures is not trivial, as described in section 6. In this paper, it is assumed that scales to measure the performance of the drugs on various criteria already exist and are specified. Furthermore, it is assumed without loss of generality that all these scores are defined in such a manner that increasing values are preferred.

Henceforth, the performance levels of drugs A and B on i th criterion are measured on these scales are represented as performance score values $v_i(a)$ and $v_i(b)$, respectively. For any criterion i , performance score $v_i(a)$ is a non-decreasing function of the attribute value $z_i(a)$; this could also be defined more generally for any criterion i as, $v_i = f(z_i)$, the function f is same for all alternatives (drugs) keeping in line with the need for the performance of the different alternatives to be measured in an objective manner.

The three different MCDA approaches mentioned are applied to this case study as below; this allows us to demonstrate the potential advantages and pitfalls of using the different MCDA modelling approaches.

5.1. VALUE MEASUREMENT MODELS

This approach is based on constructing a single overall value for each alternative in order to establish a preference order of alternatives. An alternative A is said to be preferred to B if $V(a) > V(b)$ where $V(a)$ and $V(b)$ are overall values (taking into account all n criteria) of A and B, respectively. Also, there is said to be indifference between the alternatives if $V(a) = V(b)$.

The first step in this approach is to do preference modelling i.e. constructing the performance levels of drugs A and B on all criteria, as shown in Table 2. The performance score values $v_i(a)$ and $v_i(b)$ of drugs A and B on i th criterion are also known as partial value functions. The importance of different criteria is measured using the gain associated with replacing the worst outcome by the best outcome and the weights w_i represent the relative importance of i th

criterion. The final step is to aggregate these partial value functions taking into account the relative importance of different criteria; the manner in which this is done separates different value measurement approaches.

In this paper, additive aggregation (also known as weighted sum approach) is described as it is the most common value measurement modelling approach and it is based on the following equation

$$V(a) = \sum_{i=1}^n w_i v_i(a)$$

where $V(a)$ is the overall value, w_i represents the relative importance and $v_i(a)$ represents the score of alternative A on i th criterion (usually standardised in scales of 0-1, 0-10 or 0-100), respectively.

Table 2 Performance levels of drugs

Criteria (i)	Drug A $v_i(a)$	Drug B $v_i(b)$	Weights ω_i
C/E	0.72	0.84	8
Equity (%)	0.14	0.08	1
Innovation	0.91	0.62	3
Patient compliance (%)	0.93	0.85	2
Quality of evidence	0.82	0.79	3

This approach requires some assumptions regarding the criteria and their weights; namely preferential independence of criteria and the need for the weights to satisfy the trade-off requirements. Preferential independence requires that the decision can be made using a subset of criteria, if the other criteria are the same for all alternatives irrespective of their actual values i.e. the decision can be made using only the criteria on which the alternatives differ. The weight parameters w_i also need to follow a strict trade-off condition in order to capture the concept of “importance” as well as compensating for the different measurement scales of different criteria. This is achieved by swing weights which represent the gain in overall value by going from the

worst value to best value in each criterion i.e. for any two criteria i and k , the ratio w_i/w_k is the change in $v_k(a)$ that should compensate for a unit loss on $v_i(a)$. There are a number of ways in which these swing weights can be elicited, these techniques are not discussed here as they are explained in detail in section VII and in literature.⁷

In this case study, it is assumed that the relative weights and the performance of the alternatives on different criteria have been identified using appropriate techniques and are as shown in Table 2. Using this information, the overall values of drugs A and B can be calculated as below

$$V(a) = \sum 8 * 0.72 + 1 * 0.14 + 3 * 0.91 + 2 * 0.93 + 3 * 0.82 = 12.95$$

$$V(b) = \sum 8 * 0.84 + 1 * 0.08 + 3 * 0.62 + 2 * 0.85 + 3 * 0.79 = 12.73$$

This approach is simple to use but as observed in this scenario, poor performance on a criteria (C/E) can be overcome by doing well in other criteria depending on the weights and partial value functions. Also, the strict theoretical basis of this approach means that considerable caution needs to be taken to satisfy the preferential independence of criteria and the corresponding trade-offs of swing weights.

5.2. OUTRANKING APPROACH

This principle of outranking is based on the general concept of dominance.^{45,46} If the two alternative drugs A and B are such that $v_i(a) \geq v_i(b)$ for all criteria (with strict inequality for at least one criterion) where performance of each drug on each of the “ i ” criteria are $v_i(a)$ and $v_i(b)$, then we can conclude that drug A should be preferred to drug B. In this event, drug A is said to dominate drug B. However, this rarely occurs in practice and thus the evidence needs to be evaluated in a systematic manner. More generally, drug A outranks alternative drug B if there is sufficient evidence to justify a conclusion that drug A is at least as good as drug B, taking all criteria into account.

This approach utilises outranking relation (i.e. comparing performance scores on individual criterion to see which alternative outranks the other on that criterion) on a set of alternatives focusing on pair wise comparisons and these pair wise comparisons are used to estimate the concordance and discordance indices. For drug A, concordance index is the evidence in favour of

A outranking B while the discordance index is evidence against A outranking B. Similarly, for drug B, the concordance index is the evidence in favour of B outranking A while the discordance index is evidence against B outranking A.

The first step in estimating the concordance and discordance indices is to construct a matrix of outranking relations from the individual scores on each criterion. The performance scores of drugs against the individual criteria is shown in Table 3a while the matrix of outranking relations along with the relative weights for different criteria is as shown in Table 3b. The outranking approach recognises that performance scores, $v_i(a)$ and $v_i(b)$, are imprecise measures so alternative a is preferred to alternative b only if $v_i(a)-v_i(b)$ exceeds a predefined “indifference threshold”. For example, if the threshold was 0.05, alternative drug A and drug B would be incomparable on “quality of evidence” criteria as the difference 0.03 is less than the threshold. Also, it is to be noted that the weights of different criteria do not need to follow the theoretical concept of trade-offs as required by the value measurement approach, they just represent the relative importance of different criteria.

Table 3a: Performance scores of drugs

Criteria (i)	Drug A $v_i(a)$	Drug B $v_i(b)$
C/E	0.72	0.84
Equity (%)	0.14	0.08
Innovation	0.91	0.62
Patient compliance (%)	0.93	0.85
Quality of evidence	0.82	0.79

Table 3b: Outranking relations and weights

Weights ω_i	Drug A	Drug B
10		\dot{u}
2	\dot{u}	
1	\dot{u}	
3	\dot{u}	
2	-	-

There are a number of ways to quantify concordance and discordance indices which correspond to different outranking methods. In this study, ELECTRE I⁴⁷ is used but the reader should bear in mind that there are a number of other options (ELECTRE II, III, IV, TRI^{45,46,48} and PROMETHEE,⁴⁹ GAIA⁵⁰). In ELECTRE I, the concordance index is defined as the ratio of sum

of weights in criteria where drug A is at least as good as drug B to sum of weights in all criteria i.e.

$$C(a,b) = \frac{\sum_{i \in Q(a,b)} w_i}{\sum_{i=1}^m w_i}$$

where $Q(a,b)$ is set of criteria where A is atleast as good as B. For the discordance index, a veto threshold t_i can be specified as below

$$D(a,b) = \begin{cases} 1 & \text{if } v_i(b) - v_i(a) > t_i \text{ for any } i \\ 0 & \text{otherwise} \end{cases}$$

The concordance and discordance indices are compared against the concordance (C') and discordance (D') thresholds, respectively, to estimate the outranking relation. If the concordance index (C) is greater than concordance threshold (C') and discordance index (D) is less than discordance threshold (D'), then that drug is said to outrank the other drug. If the thresholds are specified such that both drugs outrank each other, then they are said to be indifferent. In case of neither drug outranking the other, the drugs are said to be incomparable. If there are more than two alternatives (i.e. A, B, C etc) the concordance and discordance indices are estimated for each pair to build an outranking relation using the concordance and discordance thresholds, C' and D' respectively.

In the current case study, the concordance index of drug A against drug B is sum of weights in $Q(a,b)$, set of criteria where A is at least as good as B, divided by overall sum of weights i.e. $(2+1+3+2)/(10+2+1+3+2) = 8/18$. However, the decision can be vetoed by the poor performance of drug A in cost-effectiveness by specifying the veto threshold t_1 as 0.1 for C/E (drug B is better than A in C/E by 0.12 which is greater than veto threshold t_1). Similarly, the concordance index of drug B against drug A is 10/18. If the concordance threshold (C') is less than 0.56, drug B is said to outrank drug A provided the veto thresholds for other criteria are not violated as the concordance index is greater than the concordance threshold.

The advantages of this method are as below:

1. No theoretical requirement of weights as required in the value measurement models; they just convey the relative importance of the different criteria

2. Intuitive, reflects the current NICE process
3. Different levels of complexity: ELECTRE I, II, III, IV, TRI and PROMETHEE, GAIA

A potential pitfall might be that this approach might lead to incomparability if two drugs are quite similar; however, one could argue that this is appropriate for the NICE appraisal process as further deliberation might be needed to choose between the two drugs.

5.3. GOAL PROGRAMMING

This approach is based on satisficing,⁵¹ where the emphasis is on attaining satisfactory levels of performance on each criteria with preference given to the criteria in the order of importance. Goal programming approach involves mathematical formulation of the satisficing heuristic, the satisficing levels are predefined as “goals” classified according to importance, and a programming algorithm is used to identify the alternatives which satisfy the goals in the specified priority order.⁵² This is the same as solving linear programs with multiple objectives so this approach is based on linear programming (LP).

Unlike the weighted sum approach which involves developing partial value functions $v_i(a)$, the goal programming method operates directly on the attribute values, $z_i(a)$, of the alternatives on the criteria, as it is more operationally meaningful to match measurable attributes to the goals. The attribute values of alternative A corresponding to the “n” criteria are represented as $z_1(a)$, $z_2(a)$, ... $z_n(a)$ while the “goals” for each criterion are represented as g_1, g_2, \dots, g_n as shown in Table 4.

It should be noted that the goals vary in the direction of preference, which signifies the relationship between the attribute value and the goal. It could be a) maximizing, i.e. the goal represents a minimum performance level which deems to be satisfactory such as attaining at least 95% patient compliance, b) minimizing, i.e. goal represents a maximum level of performance which deems satisfactory (such as an ICER threshold of £10,000/QALY), and c) attainment, i.e. the attribute must achieve as close to the goal as possible. The difference between the attribute values and the goals are represented as goal deviations d_i^+ or d_i^- i.e. the amounts a targeted goal is exceeded or underachieved, respectively.

Goal programming involves minimising the goal deviations, taking the relative importance of goals into account. There are two main variants of goal programming,⁵³ weighted goal programming and lexicographic goal programming, which differ in the way the optimum solution is prioritised and achieved. The weighted goal programming approach minimises the unwanted deviations after assigning weights to the goal deviations according to their relative importance as shown in equation below:

$$\min D = \sum_{i=1}^n (\omega_i^+ d_i^+ + \omega_i^- d_i^-)$$

such that $f(x) - d_i^+ + d_i^- = g_i$ for $i = 1, \dots, n$ where x is the independent variable(s), $f(x)$ is the attribute value $z_i(a)$ as a function of the independent variable(s), g_i is the target value for i th criteria, d_i^+ and d_i^- represent the negative and positive deviations from this target value while ω_i^+ and ω_i^- are the respective weights attached to these deviations. The lexicographic goal programming formulation orders the goals into a number of priority levels and minimises them in a lexicographic manner i.e. deviation in a higher priority level being more important than any deviations in lower priority levels. This sequential minimisation approach minimises each priority whilst maintaining the minimal values reached by all higher priority level minimisations by adding them as explicit constraints.

Table 4 Attributes of drugs, the goals and weights against different criteria

Criteria (i)	Drug A $z_i(a)$	Drug B $z_i(b)$	Goals g_i	Weights ω_i^+	Weights ω_i^-
C/E (measured as NB)	£15,850	£25,600	£20,000	0	10
Equity (%)	0.14	0.08	0.20	0	5
Innovation	Innovative	Less Innovative	-	0	0
Patient compliance (%)	0.93	0.85	0.95	0	5
Quality of evidence	Good	Good	-	0	0

In this case study, it is assumed that patient compliance and equity is difficult to change but C/E can be improved by changing the price of the drug (akin to value based pricing). As C/E is the

only attribute that can be changed, it does not matter whether a weighted GP approach or a lexicographic GP approach is utilised. In this study, the lexicographic GP approach is used with C/E as the highest priority and all the other criteria together as the next priority. Assuming the net benefit for drug A varies according to $f(x) = 25000 - 1000x$ where x is the unit price of the drug A, the price of the drug A has to be decreased by 45% (from initial price of £9.15 to £5.00) so that the C/E goal of a net benefit of £20,000 is achieved. The NB for drug B is already above the specified goal threshold. Now that both the drugs have achieved the target C/E, the analysis can move to the next priority level which includes all the other criteria. Thus, the deviations for drug A and B are

$$D(a) = \sum 5 * 0.06 + 5 * 0.02 = 0.4$$

$$D(b) = \sum 5 * 0.12 + 5 * 0.1 = 1.1$$

Drug A performs better than drug B in terms of getting closer the equity and compliance goals, thus, it could be recommended on the condition that its price is reduced by 25% (in order to ensure drug A satisfies the C/E goal).

Another variation to this approach is to have a range of goals based on other criteria. For example, the C/E could have different thresholds based on the alternative's performance on other criteria.¹³ In practice, this could be implemented by using a higher C/E threshold if the aggregated goal deviations for other criteria are low i.e. a technology could be assigned a higher C/E threshold (as it performs better in achieving close to other goals).

This goal programming approach echoes similarities with value based pricing,^{43,44} but care needs to be taken in terms of choosing the appropriate goal programming strategy based on the definition of “value” chosen by the health organisations. Also, obviously, $f(x)$ will never be as simple as our assumption so complex cost-effectiveness models need to be built and analysed to identify the price of the drug such that the ICER is under the recommended threshold.

6. GENERIC ISSUES WITH USING MCDA

The NICE technology appraisal process already includes the identification of options (treatments, drugs, etc) to be appraised against the criteria specified in the NICE guidance documents. The decision makers are also identified by NICE apriori and they comprise the different NICE appraisal committees. Thus, the potential issues that might arise with implementing MCDA in the HTA process are provided below

- a) Appraisal specific or generic process: The MCDA process could be the same for all technology appraisals or it could be tailored to a given appraisal under consideration. For example, the functions which estimate the value of alternatives against the criteria could either be the fixed for all the appraisals or appraisal-specific functions could be built based on appraisal characteristics. Thus, appropriate care needs to be given before deciding on appraisal-specific process or a standard approach for all the appraisals.
- b) Choosing the criteria: This paper does not argue which criteria should be included in the decision process, however, for the success of MCDA the criteria should satisfy certain characteristics such as relevance, completeness, non-redundancy, understandability and feasibility. They should also be clearly defined, judgmentally independent and scalable (i.e. measurable in an objective manner).
- c) Modelling performance scores: This involves calculating the scores that reflect the value of an option's expected performance on the criteria in an objective manner. In some cases, the attribute values can substitute for performance scores. However, a linear relationship between the attribute values and performance measures is rare and an ordinal or ordered categorical scale, also known as partial value function, needs to be constructed. If a criterion is not linked with a measurable attribute, the performance scores can be derived using qualitative value scales⁵⁴ or direct rating.⁴⁰ This is not a trivial process and considerable care needs to develop these scales/performance scores.

- d) **Understanding weights:** This involves interpreting the criteria weights that measure the relative importance of one criterion compared to others. Unlike the measurement of performance values, this depends upon the type of MCDA approach used. In the weighted sum approach, the meaning of weights is clearly defined in terms of the trade-offs between the criteria and thus “swing weights” need to be estimated. As the swing weights are dependent upon the scales, care needs to be taken when using different performance scales; also, the weights need to adhere to some theoretical constraints in the weighted sum approach. There are no explicit weights in goal programming approach but the relative importance of criteria is specified using appropriate deviations from the aspiration levels. The outranking approach uses weights in the concordance index; these weights convey the relative importance but do not have to satisfy any conditions. It is imperative to ensure that the decision makers understand and interpret the meaning of these weights before they are asked to provide the numeric inputs to estimate the weights.
- e) **Modelling weights:** The relative importance of different criteria is elicited/modelled as weights; this is done after the decision makers are informed of the meaning of weights in the specific MCDA context. Swing weights, which use trade-off methods to capture both importance as well as the effect of measurement scales, are relevant for value measurement methods and goal programming. Outranking methods use just “importance weights”, there is no standard method to elicit these weights as they are difficult to interpret, but swing weights can be applied in this method as well. The weights (i.e. preferences of the individual decision makers) are captured in a workshop setting using deliberation, visual analogue scales or by direct rating of alternatives. However, discrete choice experiments (DCEs),^{55,56} are preferred in case of a number of decision makers/participants. DCEs present participants with hypothetical scenarios (choice sets), described using consistent set of criteria, as a survey; the data on participants choices is collected and statistically analysed to elicit the relative importance that decision makers place on different criteria.⁵⁷
- f) **Group dynamics:** The decision committee includes a number of individuals, thus appropriate care needs to be taken in the collection and aggregation of the individuals’ preferences (both in modelling criteria weights as well as preference scores). The method for aggregation of

individuals' preferences is dependent on whether a consensus needs to be achieved by the committee. If a consensus is necessary, the individuals in the committee need to share/compare their values in order to identify issues of conflict and achieve common ground. Otherwise, the overall value can be calculated as an average of the individual values which could be anonymous if need be.

- g) Uncertainty modelling: There are three main areas of uncertainty involved with using MCDA in HTA process, namely: uncertainty in problem structuring (i.e. choosing right MCDA model, criteria, level of detail etc); uncertainty with evidence of different alternatives and imprecision in modelling (i.e. uncertainty in performance scores, criteria weights, thresholds etc). Structural uncertainty is hard to capture and is out of the scope of this paper. Due to the interdependence of the other two uncertainties, appropriate care needs to be taken in performing uncertainty analysis.

The uncertainty in clinical and cost effectiveness as well as other evidence (usually caused by extrapolating the data from a randomised controlled trial to a general population) has a direct effect on committee members' preferences; thus, it should be ensured that they understand this uncertainty in evidence. Scenario analyses,⁵⁸ multi attribute utility theory,⁵⁹ fuzzy logic,⁶⁰ and stochastic multicriteria acceptability analysis⁶¹ can be used to capture this uncertainty. Value of information analysis has also been identified recently by the MCDA community as having the potential to evaluate the benefits of collecting additional information on uncertain evidence.⁶²

Modelling imprecision i.e. uncertainty in criteria values, weights and thresholds is also evident during the aggregation of the preferences of individuals in the decision committee; all the mean values are associated with the standard deviations which represent the variability in the preferences of the committee members. Sensitivity analysis can be performed to see the robustness of results to changes in the model parameters. Probabilistic sensitivity analysis (PSA) can also be used to capture and propagate the parameter uncertainty with the help of Monte-Carlo simulation techniques.^{63,64}

h) Other practical issues: Appropriate thought needs to be given to the practical aspects such as whether to train all the committee members in the relevant techniques of MCDA or whether to have a facilitator(s) to help use the techniques in the decision process. Also, the MCDA techniques rely on preference capturing, statistical analysis and synthesizing data which may require specialist software or programs; thus the relevant software/program requirements need to be identified. This also relates to other practical issues such as the methods of data capturing (survey sheets on paper, computer based forms, etc) and data aggregation. Data aggregation involves capturing the individual committee members' preferences and transferring them into appropriate software; this could be done in real time or in between the meetings. Appropriate caution needs to be taken for consistency in the transfer of data (from members' preferences to software); it must also be decided whether this data is visible and accessible to all the committee members. The MCDA model developed needs to be explored to ensure the robustness of key factors; this can be performed either pre-meeting, during the meeting or post-meeting depending on the availability of MCDA facilitator. Finally, the model outputs need to be visualised and incorporated into the final documentation along with the recommendations.

7. CONCLUSIONS

MCDA has been suggested for use in HTA but most of the recommendations are based on the weighted sum approach. An overview of the MCDA process is provided and is compared against NICE process; it is identified to be similar to the existing NICE appraisal process but with the addition of a formal mathematical approach to decision making. The main MCDA modelling approaches are described and their semblance to other techniques (such as PBMA, value based pricing and NICE recommended table of the summary characteristics) was identified. These MCDA approaches are applied to a hypothetical case study and their potential strengths and weaknesses are outlined. The general practical issues that might arise from using an MCDA approach in the HTA process are also described. It is suggested that appropriate care needs to be taken to address the issues identified in order to ensure the success of MCDA techniques in the appraisal process.

8. REFERENCES

1. Goldman, D., Lakdawalla, D., Philipson, T.J., Yin, W. Valuing health technologies at NICE: recommendations for improved incorporation of treatment value in HTA. *Health Econ* 2010; 19(10):1109-1116.
2. Baeten, S.A., Baltussen, R.M., Uyl-de Groot, C.A., Bridges, J., Niessen, L.W. Incorporating equity-efficiency interactions in cost-effectiveness analysis-three approaches applied to breast cancer control. *Value Health* 2010; 13(5):573-579.
3. Klein, R. Rationing in the fiscal ice age. *Health Econ Policy Law* 2010; 5(4):389-396.
4. Professor Sir Ian Kennedy. Appraising the value of innovation and other benefits: A short study for NICE. 2010; available from <http://www.nice.org.uk/media/98F/5C/KennedyStudyFinalReport.pdf>
5. Neumann, P.J., Weinstein, M.C. Legislating against use of cost-effectiveness information. *N Engl J Med* 2010; 363(16):1495-1497.
6. Department of Health. A new value-based approach to the pricing of branded medicines - a consultation. 2010; available from http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_122760
7. Belton, V., Stewart, T.J. Multiple Criteria Decision Analysis: An Integrated Approach. Kluwer Academic Publishers, 2002.
8. Dolan, P., Tsuchiya, A., Wailoo, A. NICE's citizen's council: what do we ask them, and how? *Lancet* 2003; 362(9387):918-919.
9. Dolan, P., Shaw, R., Tsuchiya, A., Williams, A. QALY maximisation and people's preferences: a methodological review of the literature. *Health Econ* 2005; 14(2):197-208.
10. Sassi, F., Archard, L., Le, G.J. Equity and the economic evaluation of healthcare. *Health Technol Assess* 2001; 5(3):1-138.
11. Williams, A.H., Cookson, R.A. Equity-efficiency trade-offs in health technology assessment. *Int J Technol Assess Health Care* 2006; 22(1):1-9.
12. Wailoo, A., Tsuchiya, A., McCabe, C. Weighting must wait: incorporating equity concerns into cost-effectiveness analysis may take longer than expected. *Pharmacoeconomics* 2009; 27(12):983-989.
13. Precision Health Economics LLC. Submission to the Kennedy Review. 2009; available from <http://www.nice.org.uk/media/B4F/7F/KennedyPHE.pdf>
14. Baltussen, R., Niessen, L. Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Eff Resour Alloc* 2006; 4:14.

15. Wilson, E.C., Peacock, S.J., Ruta, D. Priority setting in practice: what is the best way to compare costs and benefits? *Health Econ* 2009; 18(4):467-478.
16. Baltussen, R., Youngkong, S., Paolucci, F., Niessen, L. Multi-criteria decision analysis to prioritize health interventions: Capitalizing on first experiences. *Health Policy* 2010; 96(3):262-264.
17. Husereau, D., Boucher, M., Noorani, H. Priority setting for health technology assessment at CADTH. *Int J Technol Assess Health Care* 2010; 26(3):341-347.
18. Department for Communities and Local Government: London. Multicriteria analysis: a manual. 2009; available from <http://www.communities.gov.uk/documents/corporate/pdf/1132618.pdf>
19. Nutt, D.J., King, L.A., Phillips, L.D. Drug harms in the UK: a multicriteria decision analysis. *Lancet* 2010; 376(9752):1558-1565.
20. Mussen, F., Salek, S., Walker, S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf* 2007; 16 Suppl 1:S2-S15.
21. Mussen, F., Salek, S., Walker, S. Development and Application of a Benefit-Risk Assessment Model Based on Multi-Criteria Decision Analysis. John Wiley & Sons Ltd, 2008.
22. Dolan, J.G. Shared decision-making--transferring research into practice: the Analytic Hierarchy Process (AHP). *Patient Educ Couns* 2008; 73(3):418-425.
23. van Til, J.A., Renzenbrink, G.J., Dolan, J.G., Ijzerman, M.J. The use of the analytic hierarchy process to aid decision making in acquired equinovarus deformity. *Arch Phys Med Rehabil* 2008; 89(3):457-462.
24. Dolan, J.G. Medical decision making using the analytic hierarchy process: choice of initial antimicrobial therapy for acute pyelonephritis. *Med Decis Making* 1989; 9(1):51-56.
25. Hummel, J.M., Snoek, G.J., van Til, J.A., van, R.W., Ijzerman, M.J. A multicriteria decision analysis of augmentative treatment of upper limbs in persons with tetraplegia. *J Rehabil Res Dev* 2005; 42(5):635-644.
26. Mitton, C.R., Donaldson, C. Setting priorities and allocating resources in health regions: lessons from a project evaluating program budgeting and marginal analysis (PBMA). *Health Policy* 2003; 64(3):335-348.
27. Ruta, D., Mitton, C., Bate, A., Donaldson, C. Programme budgeting and marginal analysis: bridging the divide between doctors and managers. *BMJ* 2005; 330(7506):1501-1503.

28. Peacock, S.J., Richardson, J.R., Carter, R., Edwards, D. Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). *Soc Sci Med* 2007; 64(4):897-910.
29. Mitton, C.R., Donaldson, C., Waldner, H., Eagle, C. The evolution of PBMA: towards a macro-level priority setting framework for health regions. *Health Care Manag Sci* 2003; 6(4):263-269.
30. Mitton, C., Donaldson, C. Twenty-five years of programme budgeting and marginal analysis in the health sector, 1974-1999. *J Health Serv Res Policy* 2001; 6(4):239-248.
31. Posnett, J., Street, A. Programme budgeting and marginal analysis: an approach to priority setting in need of refinement. *J Health Serv Res Policy* 1996; 1(3):147-153.
32. Robinson, R. Limits to rationality: economics, economists and priority setting. *Health Policy* 1999; 49(1-2):13-26.
33. Goetghebeur, M.M., Wagner, M., Khoury, H., Levitt, R.J., Erickson, L.J., Rindress, D. Evidence and Value: Impact on DEcisionMaking--the EVIDEM framework and potential applications. *BMC Health Serv Res* 2008; 8:270.
34. Goetghebeur, M.M., Wagner, M., Khoury, H., Rindress, D., Gregoire, J.P., Deal, C. Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decision-making framework to growth hormone for Turner syndrome patients. *Cost Eff Resour Alloc* 2010; 8:4.
35. Health England. Prioritising investments in preventative health. 2009; available from <http://www.healthengland.org/>
36. Dowie, J. The future of HTA is MCDA. 2008; available from <http://knol.google.com/k/the-future-of-hta-is-mcda>
37. Phillips, L., Bana e Costa, C. Transparent prioritisation, budgeting and resource allocation with multi-criteria decision analysis and decision conferencing. *Annals of Operations Research* 2007; 154(1):51-68.
38. Schein, E. *Process Consultation Revisited: Building the Helping Relationship* (Addison-Wesley Series on Organization Development). Addison Wesley Longman, 1998.
39. Liberatore, M.J., Nydick, R.L. The analytic hierarchy process in medical and health care decision making: A literature review. *European Journal of Operational Research* 2008; 189(1):194-207.
40. Vaidya, O.S., Kumar, S. Analytic hierarchy process: An overview of applications. *European Journal of Operational Research* 2006; 169(1):1-29.

41. Professor Sir Michael Rawlins. Response to Sir Ian Kennedy's Report: Appraising the value of innovation. 2010; available from <http://www.nice.org.uk/media/A43/34/ITEM7KennedyConsultationResponse.pdf>
42. Aouni, B., Kettani, O. Goal programming model: A glorious history and a promising future. *European Journal of Operational Research* 2001; 133(2):225-231.
43. Claxton, K., Briggs, A., Buxton, M.J., Culyer, A.J., McCabe, C., Walker, S. et al. Value based pricing for NHS drugs: an opportunity not to be missed? *BMJ* 2008; 336(7638):251-254.
44. Towse, A. Value based pricing, research and development, and patient access schemes. Will the United Kingdom get it right or wrong? *Br J Clin Pharmacol* 2010; 70(3):360-366.
45. Roy, B. The Outranking Approach and the Foundations of Electre Methods. *Theory and Decision* 1991; 31(1):49-73.
46. Figueira, J., Greco, S., Ehrogott, M., Figueira, J., Mousseau, V., Roy, B. Electre Methods. *Multiple Criteria Decision Analysis: State of the Art Surveys*. 78 ed. Springer New York; 2005; 133-153.
47. Roy, B. Ranking and Choice in Pace of Multiple Points of View (Electre Method). *Revue Francaise D Informatique De Recherche Operationnelle* 1968; 2(8):57.
48. Roy, B., Vanderpooten, D. The European school of MCDA: Emergence, basic features and current works. *Journal of Multi-Criteria Decision Analysis* 1996; 5(1):22-38.
49. Brans, J.P., Vincke, P. A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). *Management Science* 1985; 31(6):647-656.
50. Brans, J.P., Mareschal, B. The PROMCALC & GAIA decision support system for multicriteria decision aid. *Decision Support Systems* 1994; 12(4-5):297-310.
51. Simon, H. *Administrative Behavior*, 4th Edition. Free Press, 1997.
52. Ignizio, J.P. A Review of Goal Programming: A Tool for Multiobjective Analysis. *The Journal of the Operational Research Society* 1978; 29(11):1109-1119.
53. Tamiz, M., Jones, D., Romero, C. Goal programming for decision making: An overview of the current state-of-the-art. *European Journal of Operational Research* 1998; 111(3):569-581.
54. Guttman, L. A Basis for Scaling Qualitative Data. *American Sociological Review* 1944; 9(2):139-150.

55. Ryan, M., Gerard, K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy* 2003; 2(1):55-64.
56. Ryan, M. Discrete choice experiments in health care. *BMJ* 2004; 328(7436):360-361.
57. Lancsar, E., Louviere, J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics* 2008; 26(8):661-677.
58. Hsia, P., Samuel, J., Gao, J., Kung, D., Toyoshima, Y., Chen, C. Formal approach to scenario analysis. *Software, IEEE* 1994; 11(2):33-41.
59. Winterfeldt, D.V., Fischer, G.W. Multi-attribute utility theory : models and assessment procedures. Kluwer, 1973.
60. Klir, G.J., Yuan, B. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall PTR, 1995.
61. Tervonen, T., Figueira, J.R. A survey on stochastic multicriteria acceptability analysis methods. *Journal of Multi-Criteria Decision Analysis* 2008; 15(1-2):1-14.
62. Linkov, I., Satterstrom, F., Steevens, J., Ferguson, E., Pleus, R. Multi-criteria decision analysis and environmental risk assessment for nanomaterials. *Journal of Nanoparticle Research* 2007; 9(4):543-554.
63. Hyde, K., Maier, H.R., Colby, C. Incorporating uncertainty in the PROMETHEE MCDA method. *Journal of Multi-Criteria Decision Analysis* 2003; 12(4-5):245-259.
64. Dorini, G., Kapelan, Z., Azapagic, A. Managing uncertainty in multiple-criteria decision making related to sustainability assessment. *Clean Technologies and Environmental Policy* 2011; 13(1):133-139.